A Geometric Review of Linear Algebra

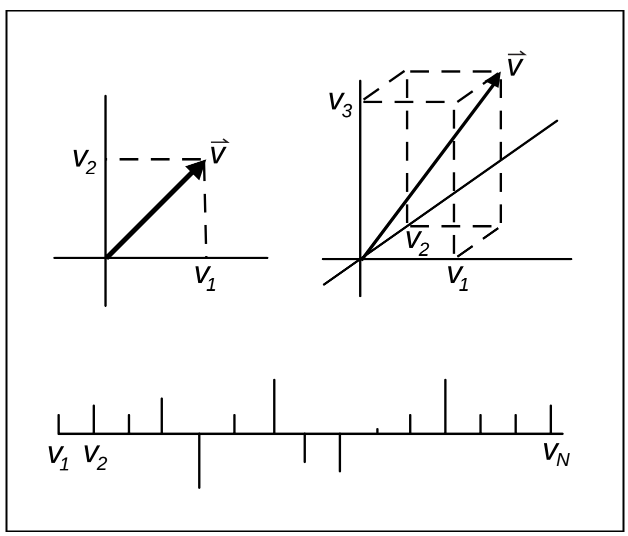
The following is a compact review of the primary concepts of linear algebra. I assume the reader is familiar with basic (i.e., high school) algebra and trigonometry. The order of presentation is unconventional, with emphasis on geometric intuition rather than mathematical formalism. For a gentler introduction, I recommend *The Appendix on Linear Algebra* from the PDP book series, by Michael Jordan. For more thorough coverage, I recommend *Linear Algebra and Its Applications* by Gilbert Strang, Academic Press, 1980.

Vectors (Finite-Dimensional)

A **vector** is an ordered collection of numbers, known as the **components** of the vector. I'll use the variable N throughout to represent the number of components, known as the **dimensionality** of the vector. We usually denote vector-valued variables with an over-arrow (e.g., $\vec{v}$). If we want to indicate the components, we often arrange them vertically (as a "column" vector):

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix},$$

Vectors of dimension 2 or 3 can be graphically depicted as arrows, with the tail at the origin and the head at the coordinate location specified by the vector components. Vectors of higher dimension can be illustrated using a "spike plot".



The **norm** (or **magnitude**) of a vector is defined as: $\|\vec{v}\| = \sqrt{\sum_n v_n^2}$, which has a value that is positive. Geometrically, this corresponds to the length of the vector (in 2 dimensions, this comes from the Pythagorean theorem). A vector containing all zero components has zero

• Author: Eero Simoncelli, Center for Neural Science, and Courant Institute of Mathematical Sciences.

• Created: 20 January 1999. Last revised: September 22, 2020.

• Send corrections or comments to eero.simoncelli@nyu.edu

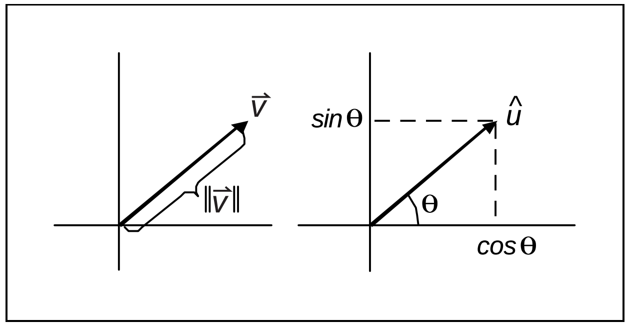
• The latest version of this document is available from: `http://www.cns.nyu.edu/~eero/NOTES/`

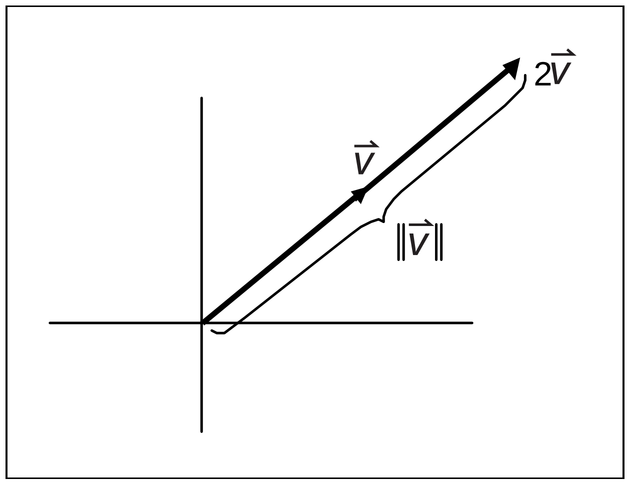
norm, and is called the **zero vector**.

A **unit vector** is a vector of length one. In two dimensions, the set of all unit vectors may be parameterized in terms of their angle relative to the horizontal axis:

$$\hat{u}(\theta) = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}.$$

Angular parameterization of unit vectors can be generalized to N dimensions, but is much messier.



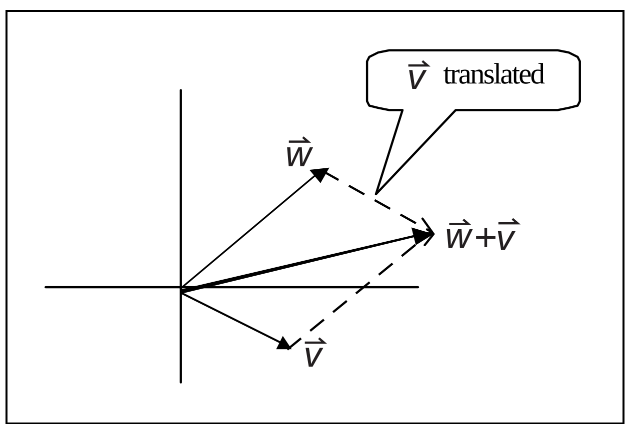


Multiplying a vector by a scalar simply changes the length of the vector by that factor. That is: $\|a\vec{v}\| = |a|\|\vec{v}\|$. Multiplying by a negative number reverses the direction. The set of vectors corresponding to all rescalings of a given vector form a straight line through the origin. Any vector (with the exception of the zero vector!) may be rescaled to have unit length by dividing by its norm: $\hat{v} = \vec{v}/\|\vec{v}\|$. Said differently, any vector may be factorized into a product of a scalar (its norm) and a unit vector (which characterizes its direction): $\vec{v} = \|\vec{v}\|\,\hat{v}$.

The **sum** of two vectors of the same dimensionality is a vector whose components are sums of the corresponding components of the summands. Specifically, $\vec{y} = \vec{w} + \vec{v}$ means that:

$$z_n = w_n + v_n, \qquad \text{for every } n$$

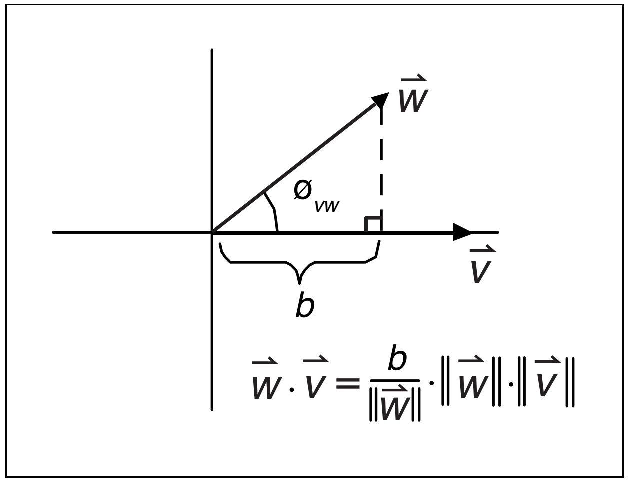
Geometrically, this corresponds to stacking the vectors head-to-foot.



The **inner product** (also called the "dot product") of two vectors is the sum of the pairwise product of components:

$$\vec{v} \cdot \vec{w} \equiv \sum_n v_n w_n.$$

Note that the result is a scalar.



This operation has an equivalent geometric definition, which allows for a more intuitive, geometric, interpretation:

$$\vec{v} \cdot \vec{w} \equiv \; \|\vec{v}\| \, \|\vec{w}\| \; \cos(\phi_{vw}),$$

where ϕ_{vw} is the angle between the two vectors. This is easy to prove in two dimensions: write the two vectors as re-scaled unit vectors, write the unit vectors in terms of their angle (as above), and then use the trig identity $\cos(\theta_w)\cos(\theta_v) + \sin(\theta_w)\sin(\theta_v) = \cos(\theta_w - \theta_v)$. This cosine is equal to $b/\|\vec{v}\|$ in the figure to the right.

Note that the inner product of two perpendicular vectors is 0, the inner product of two parallel vectors is the product of their norms, and the inner product of a vector with itself is the square of its norm.

From the definition, you should be able to convince yourself that the inner product is distributive over addition: $\vec{v} \cdot (\vec{w} + \vec{y}) = \vec{v} \cdot \vec{w} + \vec{v} \cdot \vec{y}$. And similarly, the inner product is also commutative (i.e., order doesn't matter): $\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v}$.

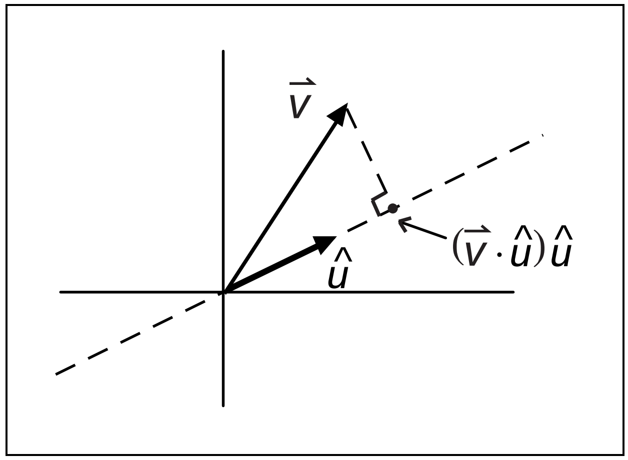
Despite this symmetry, it is often useful to interpret one of the vectors in an inner product as an operator, that is applied to an input. For example, the inner product of any vector $\vec{v}$ with the vector:

$$\vec{w} = (\, \tfrac{1}{N} \; \tfrac{1}{N} \; \tfrac{1}{N} \; \cdots \; \tfrac{1}{N} \,)$$

gives the average of the components of $\vec{v}$. The inner product of vector $\vec{v}$ with the vector

$$\vec{w} = (\, 1 \; 0 \; 0 \; \cdots \; 0 \,)$$

is the first component, v_1.



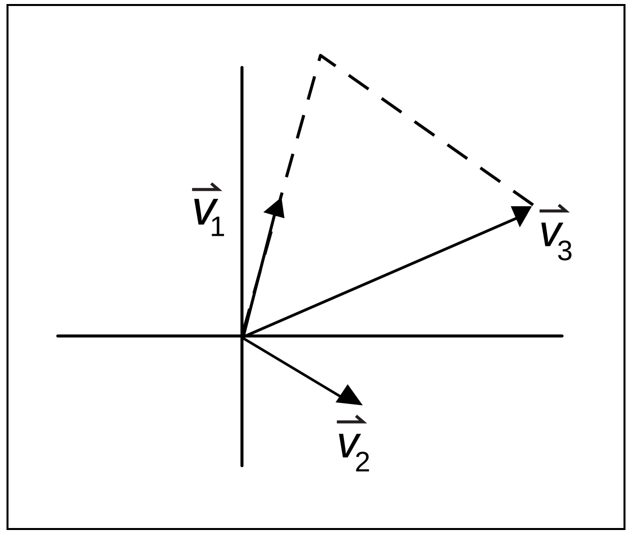
The inner product of a vector $\vec{v}$ with a unit vector $\hat{u}$ has a particularly useful geometric interpretation. The cosine equation implies that the inner product is the length of the component of $\vec{v}$ lying along the line in the direction of $\hat{u}$. This component, which is written as $(\vec{v} \cdot \hat{u})\hat{u}$, is referred to as the **projection** of the vector onto the line. The difference (or residual) vector, $\vec{v} - (\vec{v} \cdot \hat{u})\hat{u}$, is the component of $\vec{v}$ perpendicular to the line. Note that the residual vector is always perpendicular to the projection vector, and that their sum is $\vec{v}$ [prove].

Vector Spaces

Vectors live in **vector spaces**. Formally, a vector space is just a collection of vectors that is closed under linear combination. That is, if the two vectors $\{\vec{v}, \vec{w}\}$ are in the space, then the vector $a\vec{v} + b\vec{w}$ (with a and b any scalars) must also be in the space. All vector spaces include the zero vector (since multiplying any vector by the scalar zero gives you the zero vector). A **subspace** is a vector space lying within another vector space (think of a plane, slicing through the 3D world that we inhabit).

This definition is somewhat abstract, as it implies that we construct vector spaces by starting with a few vectors and "filling out" the rest of the space through linear combination. But we have been assuming an implicit vector space all along: the space of all N-vectors (denoted $\mathbf{R}^N$)

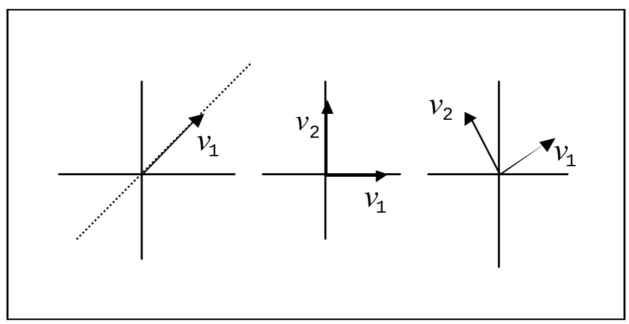
is clearly a vector space [verify].



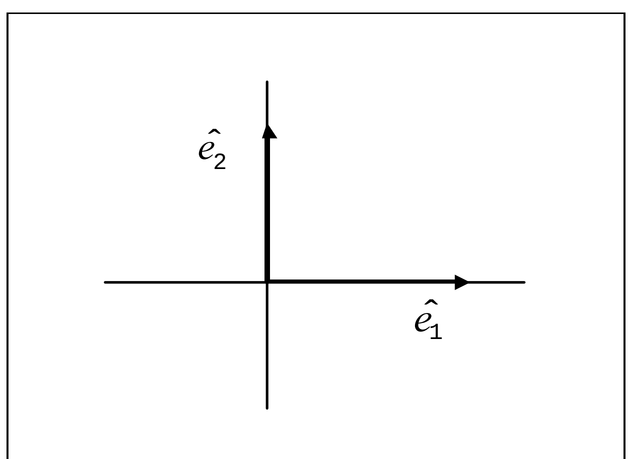
Working backwards, a set of vectors is said to **span** a vector space if one can write any vector in the vector space as a linear combination of the set. A spanning set can be redundant: For example, if two of the vectors are identical, or are scaled copies of each other. This redundancy is formalized by defining **linear independence**. A set of vectors $\{\vec{v}_1, \vec{v}_2, \ldots \vec{v}_M\}$ is linearly independent if (and only if) the only solution to the equation

$$\sum_n \alpha_n \vec{v}_n = 0$$

is $\alpha_n = 0$ (for all n).



A **basis** for a vector space is a linearly independent spanning set. For example, consider the plane of this page. One vector is not enough to span the plane: Scalar multiples of this vector will trace out a line (which is a subspace), but cannot "get off the line" to cover the rest of the plane. But two vectors are sufficient to span the entire plane. Bases are not unique: any two vectors will do, as long as they don't lie along the same line. Three vectors are redundant: one can always be written as a linear combination of the other two. In general, the vector space R^N requires a basis of size N.



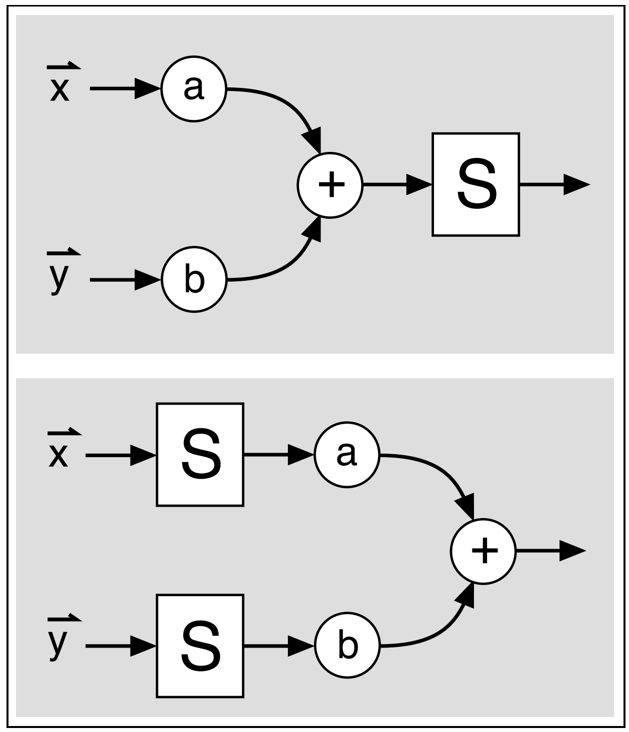
Geometrically, the basis vectors define a set of coordinate axes for the space (although they need not be perpendicular). The **standard basis** is the set of unit vectors that lie along the axes of the space:

$$\hat{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; \hat{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; \ldots \; \hat{e}_N = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Linear Systems & Matrices

A **linear system** $\mathcal{S}$ transforms vectors in one vector space into those of another vector space, in such a way that it obeys the principle of **superposition**:

$$\mathcal{S}\{a\vec{v} + b\vec{w}\} = a\mathcal{S}\{\vec{v}\} + b\mathcal{S}\{\vec{w}\}.$$



That is, the system "response" to any linear combination of vectors is equal to that same linear combination of the response to each of the vectors alone. Linear systems are useful because they are very well understood (in particular, there are powerful tools for characterizing, analyzing and designing them), and because they provide a reasonable description of many physical systems.

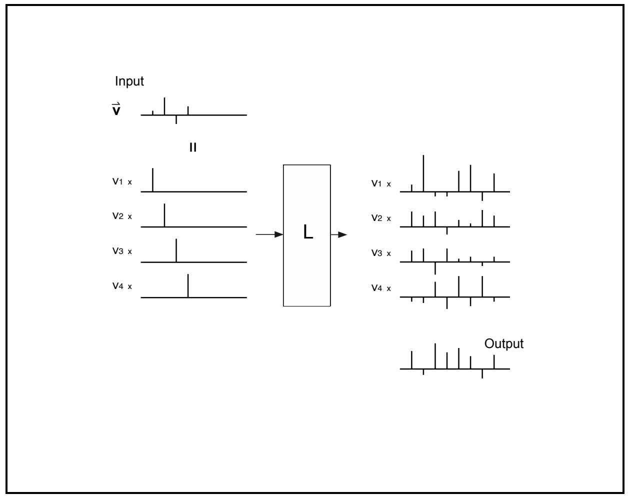
The parallel definitions of vector space and linear system allow us to make a strong statement about characterizing linear systems. First, write an arbitrary input vector $\vec{v}$ in terms of the standard basis:

$$\vec{v} = v_1\hat{e}_1 + v_2\hat{e}_2 + \ldots + v_n\hat{e}_N$$

Using the linearity of system $\mathcal{S}$, we write:

$$\begin{aligned} \mathcal{S}\{\vec{v}\} &= \mathcal{S}\{v_1\hat{e}_1 + v_2\hat{e}_2 + \ldots + v_n\hat{e}_N\} \\ &= v_1\mathcal{S}\{\hat{e}_1\} + v_2\mathcal{S}\{\hat{e}_2\} + \ldots + v_n\mathcal{S}\{\hat{e}_N\}. \end{aligned}$$

That is, the response is a linear combination of the responses to each of the standard basis vectors. Each of these responses is a vector. Since this holds for any input vector $\vec{v}$, the system is fully characterized by this set of response vectors.

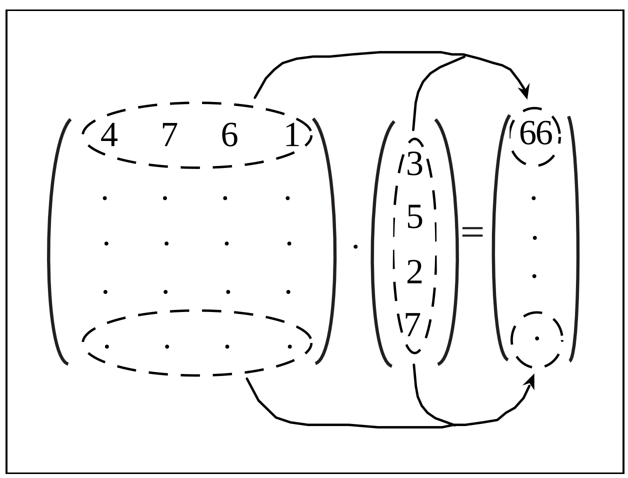


We can gather the column vectors corresponding to the responses to each axis vector into a table of numbers that we call a **matrix**. This matrix is a complete representation of the associated linear system: any response is just a linear combination (weighted sum) of the columns of the matrix. Label the elements of the matrix S_{nm} with n indicating the row and m the column. The response of the linear system to an input vector $\vec{v}$ has components

$$w_n = \sum_m S_{nm} v_m$$

The summation is over the columns of the matrix. For short, we write $\vec{w} = S\vec{v}$, and refer to this as *multiplication* of the matrix by the input vector. The operation is only defined when the number of columns in the matrix matches the dimensionality of the input vector.

An alternative interpretation of the matrix multiplication is that each component of the output vector is an inner product of the corresponding *row* of the matrix with the input vector [convince yourself that this is true].



Like vectors, matrices can be multiplied by a scalar and added (element by element). In addition, the **transpose** is another matrix with rows and columns swapped: $(S^T)_{nm} = S_{mn}$. A **symmetric** matrix is a square matrix that is equal to its transpose.

The sequential application of two linear systems is a linear system. We define the matrix associated with the full system as the **matrix product** of the two subsystem matrices. Let A

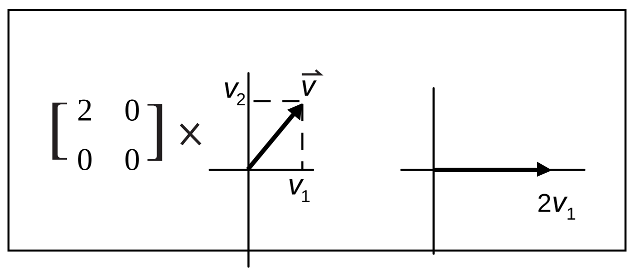
and B be the two subsystem matrices. By following the transformation of the standard basis as it passes through A and then B, we can get a definition for the product matrix C:

$$C_{nk} = \sum_m B_{nm} A_{mk}$$

The columns of the product matrix are just the application of matrix B to the columns of A. Since it can be thought of as a collection of inner products, matrix multiplication is distributive over addition. It is also associative: $A(BC) = (AB)C$. But be careful: It is generally *not* commutative.

Now consider two special classes of matrix. A **diagonal matrix** is one for which only elements along the diagonal can be non-zero. These matrices operate on vector spaces by stretching or compressing axes: the nth axis of the space is stretched or compressed by an amount specified by the nth diagonal element, S_{nn}. The product of two diagonal matrices is diagonal. If the matrix is square, and the diagonal elements are all one, the matrix does nothing. This matrix is called the **identity**, denoted I.

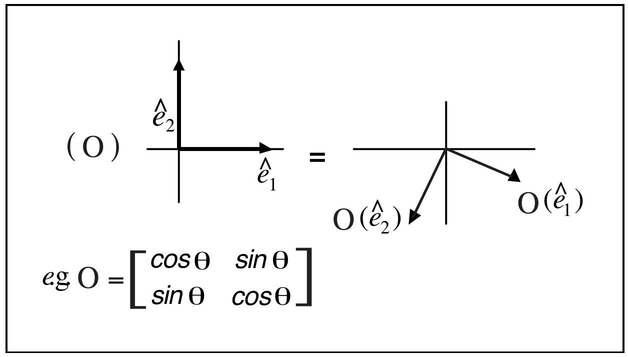
If an element of the diagonal is zero, then the associated axis is annihilated. The set of vectors that are annihilated by the matrix form a vector space [prove], which is called the **row nullspace**, or simply the **nullspace** of the matrix.



Another implication of a zero diagonal element is that the matrix cannot "reach" the entire output space, but only a proper subspace. This space is called the **column space** of the matrix, since it is spanned by the matrix columns. The **rank** of a matrix is just the dimensionality of the column space. A matrix is said to have **full rank** if its rank is equal to the smaller of its two

dimensions.

An **orthogonal** matrix is a square matrix, for which every column is a unit vector, and every pair of columns is orthogonal. This means that the transpose of the matrix multiplied by itself gives the identity matrix: $O^T O = I$ [prove]. This means that applying an orthogonal matrix to a vector does not change its length: $||O\vec{v}||^2 = (O\vec{v})^T(O\vec{v}) = \vec{v}^T O^T O\vec{v} = \vec{v}^T\vec{v} = ||\vec{v}||^2$. Similarly, applying an orthogonal matrix to two vectors does not change the angle between them [prove]. Since the columns of a matrix describe the response of the system to the standard basis, an orthogonal matrix maps the standard basis onto a new set of N orthogonal axes, which form an alternative basis for the space. Putting this all together, we can think of orthognal matrices as performing a *generalized* rotation: a rigid physical rotation of the space and possibly negation of some axes. Note that the product of two orthogonal matrices is also orthogonal.



Linear Systems of Equations

The classic motivation for the study of linear algebra is the solution of sets of linear equations such as

$$\begin{aligned}
a_{11}v_1 + a_{12}v_2 + \ldots + a_{1N}v_N &= b_1 \\
a_{21}v_1 + a_{22}v_2 + \ldots + a_{2N}v_N &= b_2 \\
&\vdots \\
a_{M1}v_1 + a_{M2}v_2 + \ldots + a_{MN}v_N &= b_M
\end{aligned}$$

If we put the variables v_n and the constants b_m into column vectors, and the constants a_{nm} into a matrix A, these equations may be written more compactly: $A\vec{v} = \vec{b}$. Furthermore, we may now use the tools of linear algebra to determine if there is a solution for $\vec{v}$.

Inverses

What if we want to invert the action of a matrix? That is: given any output vector $\vec{w}$, find the unique vector $\vec{v}$ for which $\mathcal{S}\vec{v} = \vec{w}$. If the matrix has a (nonzero) nullspace, then this is impossible. In particular, all vectors in the nullspace are mapped to 0, so given output $\vec{w} = 0$

we cannot determine from which vector in the nullspace we started. More generally, a vector $\vec{v}$ satisfying $\mathcal{S}\vec{v} = \vec{w}$ is not unique, since $(\vec{v} + \vec{n})$ also satisfies the expression (where $\vec{n}$ is chosen as any vector in the nullspace).

If a matrix has a zero nullspace, then it can be inverted. The inverse operation has an associated matrix, denoted S^{-1}. Since applying the matrix and its inverse in succession restore the original vector, the matrix product of the inverse and the matrix should be the identity: $S^{-1}S = I$.

For example, the inverse of a square diagonal matrix is another diagonal matrix. Since the original matrix stretches and compresses axes, its inverse must undo these actions by multiplying each axis by the scalar inverse of the original stretch factor. Specifically, the inverse is a diagonal matrix with elements $S^{-1}_{nn} = 1/S_{nn}$. Note that this cannot be done for diagonal elements equal to zero.

Orthogonal matrices also have simple inverses. Because of the orthogonality of the matrix columns, the transpose of an orthogonal matrix is its inverse: $O^T O = I$. Since O corresponds to a generalized rotation of the space, O^T must correspond to a generalized rotation in the opposite direction.

Finally, we've neglected non-square matrices. First, consider "short and fat" matrices, which project a vector space onto a space of smaller dimensionality. These cannot be inverted. Suppose the inverse matrix S^{-1} were to exist. Apply it to the standard basis vectors. This produces a set of vectors back in the input space of S. The span of these vectors is the column space of S^{-1}. But this space has dimensionality equal to that of the output space of S, which we assumed was smaller than that of the input space of S. Thus, our inverse cannot "reach" all the vectors in the input space of S.

A "tall and skinny" matrix embeds a vector space into one of higher dimensionality. If the matrix is full rank, this operation is invertible.

Singular Value Decomposition

The **singular value decomposition** (SVD) is a standard form for representing a matrix. It is often taught near the end of a one-semester graduate course on linear algebra (or not at all), probably because the proof is fairly involved. This is a shame, because it is one of the most fundamental and beautiful results (and extremely useful as well).

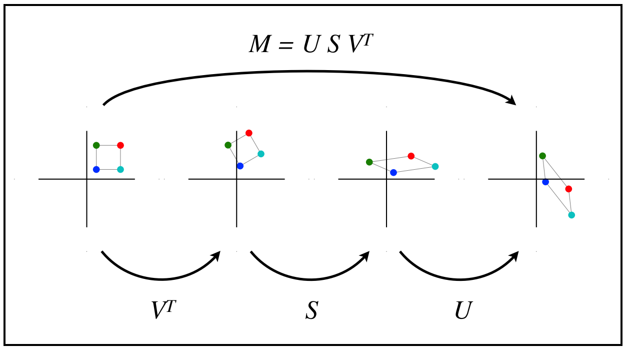
Basic result (stated without proof): *any* matrix M may be decomposed into a product of three matrices:

$$M = USV^T$$

such that U and V are orthogonal and S is diagonal with positive entries. The matrix S always

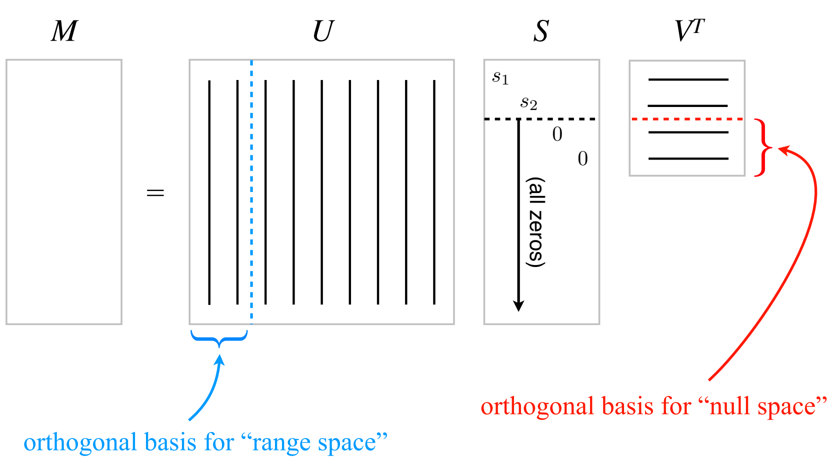
has the same dimensions as M, and the diagonal elements of S are called the **singular values**.

The advantage of this decomposition is that it describes the action of M in terms of easily understood pieces: a (generalized) rotation, scaling of the axes, and another rotation. Given this decomposition, we can directly answer many important questions regarding the matrix.



Using the diagram below, convince yourself that the following must hold:

- the nullspace of M is spanned by those columns of V associated with zero (or non-existent) singular values. Even better, these columns provide an orthogonal basis for the nullspace.
- the column space (also known as the *range space* or *target space*) of M corresponds to columns of U associated with non-zero singular values. Again, these columns provide an orthogonal basis.
- the matrix M is invertible if (and only if) the number of nonzero singular values is equal to the number of columns of M (i.e., the dimensionality of the input space).



Some "fine print"... The SVD always exists, but may be non-unique in the following ways:

1. One can permute the columns of U, diagonal elements of S, and columns of V (as long as they're all permuted the same way).
2. One can negate corresponding columns of of U and V.
3. Columns of U/V with equal corresponding singular values may be orthogonally transformed. The same transformation must be used on columns of both U and V.